Tangible reduction in learning sample complexity with large classical samples and small quantum system

Wooyeong Song[1,7] · Marcin Wieśniak[2,3] · Nana Liu[4,8,9] ·
Marcin Pawłowski[3] · Jinhyoung Lee[1] · Jaewan Kim[5] · Jeongho Bang[6]

Abstract

Quantum computation requires large classical datasets to be embedded into quantum states in order to exploit quantum parallelism. However, this embedding requires considerable resources in general. It would therefore be desirable to avoid it, if possible, for noisy intermediate-scale quantum (NISQ) implementation. Accordingly, we consider a classical-quantum hybrid architecture, which allows large classical input data, with a relatively small-scale quantum system. This hybrid architecture is used to implement a sampling oracle. It is shown that in the presence of noise in the hybrid oracle, the effects of internal noise can cancel each other out and thereby improve the query success rate. It is also shown that such an immunity of the hybrid oracle to noise directly and tangibly reduces the sample complexity in the framework of computational learning theory. This NISQ-compatible learning advantage is attributed to the oracle's ability to handle large input features.

Keywords Quantum machine learning · Probably Approximately Correct (PAC) learning · Classical-quantum hybrid query · Sample complexity

1 Introduction

Many celebrated quantum algorithms have shown promise for the quantum computational speedup [1–3]. However, apart from requiring considerable computational resources needed for the main calculation, many of them involve high costs for

---

✉ Jinhyoung Lee
   hyoung@hanyang.ac.kr

✉ Jaewan Kim
   jaewan@kias.re.kr

✉ Jeongho Bang
   jbang@etri.re.kr

Extended author information available on the last page of the article

introducing "big" classical data into quantum states, for example, by accessing a useful quantum gadget, called quantum random-access memory (QRAM). An efficient QRAM algorithm, named bucket brigade algorithm, has been developed [4,5]. Nevertheless, despite its promise for usefulness, it has some caveats related to physical resources and its ability to correct errors. This has been argued by S. Aaronson in Ref. [6]. Thus, it is still unclear whether the bucket-brigade QRAM can be a prototype. This issue is well known in quantum computation and quantum machine learning (QML) (see Ref. [7] and Chap. 5 of Ref. [8] for additional details on this issue). Therefore, researchers in the field of quantum computation and/or QML are unlikely to decisively state that *tangible* quantum (learning) speedups can be achieved with noisy intermediate-scale quantum (NISQ) technologies [9,10], when a large superposition of data is required. In this context, it is now highly desirable to show the possibility of achieving NISQ technology-based speedups, namely without using excessively large data superposition or, equivalently, without accessing the bucket-brigade QRAM.

Toward this end, one promising approach is to consider a classical-quantum hybrid architecture and identify the optimal interplay between classical and quantum strategies. Such approaches have received increasing attention, an example being the consideration of variational formulations [11–15]. Here, we devise an intriguing type of hybrid architecture, in which the input data remain classical but a small-scale quantum system is employed. Such an architecture differs from those of other classical-quantum hybridization; however, it will render the task suitable for the NISQ implementation without requiring an excessively large superposition of the inputs, or equivalently, the bucket-brigade QRAM [16–18]. The motivation and background of this study are similar to those of the recent works in Refs. [19,20].

We apply our hybridization to a classification task, a fundamental problem in computation and machine learning. For this, we employ a classical-quantum hybrid oracle designed on the basis of our main idea and assume that the oracle generates noisy samples due to errors resulting from the use of erroneous (internal) quantum gates. A model with such erroneous gates is often referred to as a noisy query model and is integrated in realistic models [21,22]. We demonstrate both analytically and numerically that our hybrid oracle can exhibit a high probability of success for queries. This advantage is attributed to the high capability of the oracle to explore a wider space of solutions, and it naturally leads to a quantum learning advantage, namely, a reduction in the sample complexity, in a computational learning framework [23,24]. Our hybridization architecture is applicable in the relevant context of NISQ machine learning [15].

2 Classification with noisy samples

Classification is a fundamental computational problem that is defined as follows [25–27]. Consider a Boolean function $h^\star$ that maps $\mathbf{x} = x_1 x_2 \ldots x_n$ ($x_j \in \{0, 1\}$ $\forall j = 1, \ldots, n$) to $h^\star(\mathbf{x}) \in \{0, 1\}$.[1] Here, an algorithm (or a learner) can invoke an oracle to prepare a (finite) set of samples; given the inputs $\mathbf{x}$, the oracle returns the

---

[1] Here, we consider a binary classification, i.e., mapping $\{0, 1\}^n \to \{0, 1\}$.

corresponding outputs $h^\star(\mathbf{x})$ and generates a set of samples $T = \{(\mathbf{x}, h^\star(\mathbf{x}))\}$. The task is to identify a hypothesis h close to $h^\star$ from among a set, say H, of candidates while minimizing the number of invocations of the oracle, or equivalently, the size of samples $|T|$. The oracle can be implemented by using the general form of the Boolean function [28]

$$h^\star(\mathbf{x}) = a_0 \oplus a_1 x_1 \oplus a_2 x_2 \oplus a_3 x_1 x_2 \oplus \cdots \oplus a_{2^n-1} x_1 x_2 \ldots x_n, \tag{1}$$

where $a_k \in \{0, 1\}$ ($k = 0, 1, \ldots, 2^n - 1$) are known as the Reed-Muller coefficients. In other words, given $\mathbf{x} = x_1 x_2 \cdots x_n$, the oracle outputs, which could be either ideal or noisy. In this study, we are interested in noisy outputs, namely, the case where the oracle provides invalid outputs, resulting in a noisy sample set $\tilde{T}$. Note that the classification problem for noisy samples has widely been studied in computational learning theory since real-world data are not clean, which makes the computation much harder [29,30].

Here, we can introduce a corresponding error model of the generation of $\tilde{T}$. For this, each coefficient a_k is changed such that $a_k \rightarrow a_k \oplus 1$ with a certain probability $\eta_k \leq \frac{1}{2}$, and from the changed coefficients, noisy samples $(\mathbf{x}, h^\star(\mathbf{x}) \oplus e)$, where e is to be 0 (or 1) with probability (say) $P(\mathbf{x})$ (or $1 - P(\mathbf{x})$), can be obtained. This situation is often referred to as classification error in a noisy query model of computation [31]. Such a model is useful since Eq. (1) is a general form and each a_k can be realized by a Toffoli gate conditioned by $\mathbf{x}$ channels in a circuit [32,33] (this is shown later). In this case, the oracle's invalid outputs are provided by flipping (i.e., $0 \rightleftharpoons 1$) the bit signal with probability η_k before or after applying the kth Toffoli gate, namely a_k. This error model can indicate a situation where the probability $P(\mathbf{x})$ of a successful query decreases as the problem size n increases, and it is well suited for determining the effect of noise in complexity-theoretic studies [16].
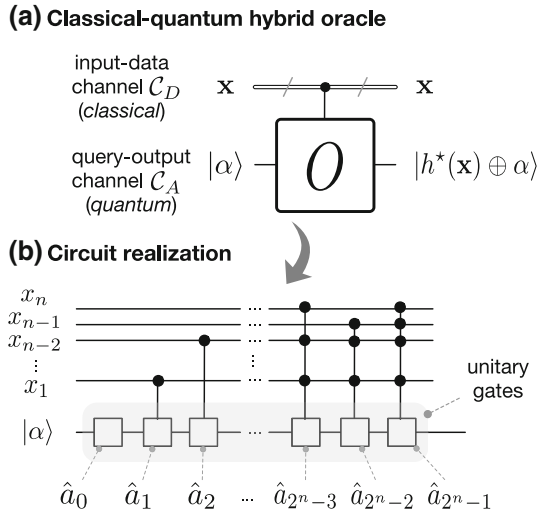
3 Classical-quantum hybrid oracle

In classical studies, the oracle handles classical inputs and outputs. By contrast, (fully) quantum approaches begin by changing classical inputs into the corresponding quantum states, such as $\mathbf{x} \rightarrow |\mathbf{x}\rangle$, and the oracle performs the mapping

$$\sum_{\mathbf{x}} |\mathbf{x}, \alpha\rangle \rightarrow \sum_{\mathbf{x}} |\mathbf{x}, h^\star(\mathbf{x}) \oplus e\rangle, \tag{2}$$

in which the quantum parallelism resulting from the superposition of the samples is utilized.[2] This oracle enables us to enjoy a quantum computational advantage in a specific situation [27]. However, such a process involves considerable cost and can even offset the advantages gained, as described at the beginning of this paper.

---

[2] Here we should clarify that such an oracle is employed for *data-sampling*, which it differs from those employed, for example, in the context of so-called amplitude amplification, where the oracle *marks* the relative phase on a single data state among superposed ones. In amplitude amplification studies, the primary objective is to reduce the number of *iterations* of the phase-marking oracle by using other incorporating modules [34].

Fig. 1 a Schematic of our hybrid oracle. The oracle consists of two different input/output (I/O) channel types, one for input classical data $\mathbf{x} = x_1 x_2 \cdots x_n$ ($x_j \in \{0, 1\} \, \forall j = 1, \ldots, n$), where n can be very large, and the other for a single qubit, which produces the query-output states $\left| h^\star(\mathbf{x}) \right\rangle \in \{0, 1\}$. **b** Circuit realization of the oracle. This oracle applies 2^n unitary gates $\hat{a}_k \in \{\hat{\sigma}_z, i\hat{\sigma}_y\}$ ($k = 0, 1, \ldots, 2^n - 1$) onto the qubit system, conditioned on the values of the classical bits x_j in $\mathbf{x}$. In a fully classical case, these gates are either identity or logical-not. Here, we set $\alpha = 0$ for convenience

(a) Classical-quantum hybrid oracle

input-data channel $\mathcal{C}_D$ (*classical*)

query-output channel $\mathcal{C}_A$ (*quantum*)

(b) Circuit realization

unitary gates

In line with our motivation and goal, we thus consider a classical-quantum hybrid oracle that allows classical inputs $\mathbf{x}$ and a small (a single qubit in our case) quantum system for the output (see Fig. 1a). More specifically, the oracle performs the mapping

$$(\mathbf{x}, |\alpha\rangle) \rightarrow (\mathbf{x}, |\psi_{\text{out}}(\mathbf{x})\rangle), \tag{3}$$

where $|\alpha\rangle$ is an arbitrary fiducial state and the output state $|\psi_{\text{out}}(\mathbf{x})\rangle$ has the form [18,22]

$$|\psi_{\text{out}}(\mathbf{x})\rangle = \sqrt{P(\mathbf{x})} \left| h^\star(\mathbf{x}) \right\rangle + \sqrt{Q(\mathbf{x})} \left| h^\star(\mathbf{x}) \oplus 1 \right\rangle, \tag{4}$$

where $Q(\mathbf{x}) = 1 - P(\mathbf{x})$. Then, we can obtain a valid (or invalid) sample with probability $P(\mathbf{x})$ (or $Q(\mathbf{x})$), which is consistent with the error model described above. Note that $\mathbf{x}$ remains unaltered during and after the mapping.

This hybrid oracle can be realized by using the circuit shown in Fig. 1b. The circuit contains 2^n gates acting on the ancilla qubit, namely the single-qubit gate $\hat{a}_0$ and $2^n - 1$ of gates $\hat{a}_k$ ($k = 1, 2, \ldots, 2^n - 1$) conditioned on the classical bit values $x_1, x_2, \ldots, x_n$ in $\mathbf{x}$. The gates $\hat{a}_k$ are

$$\hat{a}_k \in \left\{ \hat{\sigma}_z, i\hat{\sigma}_y \right\}, \quad \text{for all } k = 0, 1, \ldots, 2^n - 1, \tag{5}$$

where $\hat{\sigma}_x$, $\hat{\sigma}_y$, and $\hat{\sigma}_z$ are the Pauli operators. Such a circuit realization is consistent with that described in the previous section and Eq. (1). Actually, each coefficient a_k has a corresponding gate operation $\hat{a}_k$. More specifically, $a_k = 0$ means that $\hat{a}_k$ leaves the bit signal unchanged (identity) and $a_k = 1$ indicates that $\hat{a}_k$ flips the bit signal (logical-not). The oracle is thus characterized by a fixed set of $\hat{a}_k$'s. The gates $\hat{a}_k$ and their operation are not opened, and the task is to identify all of them.

We can consider an error model consistently where the incorrect oracle answer $h^\star(\mathbf{x}) \oplus 1$ arises from the systematic errors in the gates $\hat{a}_k$. More specifically, the error can be described by $|\psi_k(\mathbf{x})\rangle \rightarrow |\psi'_k(\mathbf{x})\rangle = \hat{\epsilon}_k |\psi_k(\mathbf{x})\rangle$, where $|\psi_k(\mathbf{x})\rangle$ is the state passing through the gate $\hat{a}_k$. Here, $\hat{\epsilon}_k$ is a bit-flip operator defined as $\hat{\epsilon}_k = \sqrt{1 - \eta_k}\,\hat{\mathbb{1}} \pm i\,\sqrt{\eta_k}\,\hat{\sigma}_x$. In addition, we should consider another type of quantum error, the phase-flip error. This is a challenge for our hybrid oracle because in general, the phase-flip does not occur in classical gates. Thus, the prospect of quantum advantage (to be proved) can be claimed more confidently. The aforementioned error model is realistic, for example, for ion-trap or superconducting qubits, where the systematic errors are caused by imperfect control pulses [35].

4 Analysis

We now analyze the query success probability $P_{C,Q}$ in Eq. (4). Here, the subscripts Q and C refer to when the ancilla system in our hybrid oracle is quantum or classical, respectively. First, let us define a set $\Omega_{\mathbf{x}} = \{0, l_1, l_2, \ldots, l_{\kappa-1}\}$ whose elements are the indices of the gates $\hat{a}_k$ that are "activated" (i.e., when the corresponding classical control bit $x_k = 1$). The number of activated gates is given by $\kappa = 2^{\omega(\mathbf{x})}$, where $\omega(\mathbf{x})$ denotes the Hamming weight of $\mathbf{x} = x_1 x_2 \cdots x_n$, namely, the number of x_j's with a value of 1 for $j \in [1, n]$. Then, $P_{C,Q}$ can be written in terms of ω. In the classical case, $P_C(\omega)$ can be estimated as

$$P_C(\omega) \simeq \sum_j^{\kappa/2} \binom{\kappa}{2j} (1 - \overline{\eta})^{2j}\,\overline{\eta}^{\kappa-2j} \simeq \frac{1}{2}\left(1 + e^{-\frac{2^\omega}{c}}\right), \tag{6}$$

where $\overline{\eta}$ is the average error probability given by $\overline{\eta} = \frac{1}{|\Omega_{\mathbf{x}}|}\sum_{k\in\Omega_{\mathbf{x}}}\eta_k$. The variance of the error probability, given by $\Delta_\eta^2 = \frac{1}{|\Omega_{\mathbf{x}}|}\sum_{k\in\Omega_{\mathbf{x}}}\eta_k^2 - \overline{\eta}^2$, is assumed to be small. The factor c is defined as

$$c = -\frac{1}{\ln(1 - 2\overline{\eta})} \simeq (2\overline{\eta})^{-1}, \tag{7}$$

and is termed characteristic constant. Here, it is assumed that $O(\overline{\eta}^2) \rightarrow 0$. From Eq. (6), c can be interpreted as the characteristic number, say κ, of steps in the gate operations allowed before the oracle begins to give completely random outputs which cannot be used for learning.

In the case of our hybrid architecture, the success probability $P_Q(\omega)$ is expressed as

$$P_Q(\omega) = \left|\langle h^\star(\mathbf{x})|\,\hat{\epsilon}_{l_{\kappa-1}}\hat{a}_{l_{\kappa-1}}\cdots\hat{\epsilon}_{l_1}\hat{a}_{l_1}\hat{\epsilon}_0\hat{a}_0\,|\alpha\rangle\right|^2. \tag{8}$$

Here, using

$$\{\hat{\sigma}_x, \hat{a}_k\}_+ = \hat{\sigma}_x\hat{a}_k + \hat{a}_k\hat{\sigma}_x = 0, \tag{9}$$

we can show that $P_Q(\omega)$ becomes unity in the limit $\Delta_\eta \rightarrow 0$. Thus, as long as the gate errors are regular, namely, $\eta_k = \overline{\eta}$ ($\forall k \in \Omega_{\mathbf{x}}$) [22], our hybrid oracle makes no

mistakes. Evidently, our gates $\hat{a}_k$ in Eq. (5) satisfy the anticommutation relation in Eq. (9), resulting in the amplitudes associated with gate errors *canceling out* through destructive interference.

However, it is impractical to realize such a perfect oracle that makes no mistakes, since in a realistic situation, it is difficult to meet the condition $\Delta_\eta = 0$. Furthermore, we should consider the phase-flip; this is crucial to generate a successful query output since the amplitudes changed by the gate errors would interfere in a disorderly manner because of the phase-flip. In fact, such a feature is often encountered in many physics models, for example, when dealing with localization problems [36]. Consequently, $P_Q(\omega)$ has a form analogous to that of Eq. (6). The characteristic constant c is replaced with an "effective" one $c_{\text{eff}} \simeq (2\overline{\eta}_{\text{eff}})^{-1}$, where again, $O(\overline{\eta}^2) \to 0$. Here, $\overline{\eta}_{\text{eff}}$ is defined in terms of the effective average error of the $\hat{a}_k$'s. Notably, it is considerably smaller than $\overline{\eta}$, since $\overline{\eta}_{\text{eff}}$ originates from the errors remaining after destructive interference. Surprisingly, this feature, namely,

$$\overline{\eta}_{\text{eff}} \leq \overline{\eta} \text{ or equivalently } c_{\text{eff}} \geq c, \tag{10}$$

does not depend on $\overline{\eta}$ but rather on Δ_η. For a more detailed analysis, see "Appendix A".

On the basis of this feature, we show a quantum advantage. We begin with the average Hamming weight $\overline{\omega} = \frac{n}{2}$ for a given number n of input bit strings. Then, on average, our hybrid oracle is useful up to the input bit-string length $n = 2\log_2 c_{\text{eff}}$, while $n = 2\log_2 c$ is the upper limit in the purely classical case. Hence, if $c_{\text{eff}} \geq c$, our hybrid oracle can be useful for larger bit-string inputs; this condition also implies the expansion of the space of legitimate samples that can be explored by the erroneous gates, which ranges from $O(e^{(2\overline{\eta})^{-2}\ln 2})$ to $O(e^{\gamma^2(2\overline{\eta})^{-2}\ln 2})$, where $\gamma = \frac{c_{\text{eff}}}{c}$ is larger than 1.

In addition to our theoretical analysis, we present numerical simulations in which $P_{C,Q}(\omega)$ are evaluated by counting a large number ($\simeq 10^5$) of queries for each given number of $\omega(\mathbf{x})$. Here, the identified data of $P_{C,Q}(\omega)$ are averaged over the trials ($\simeq 10^3$) again. To simulate a more realistic scenario, it is assumed that η_k's are drawn from a normal distribution $\mathcal{N}(\overline{\eta}, \Delta_\eta)$. We also assume that the ancilla qubit undergoes the phase-flip errors on the gates $\hat{a}_k$ with probability $\chi_k \leq \frac{1}{2}$, drawn from $\mathcal{N}(\overline{\chi}, \Delta_\chi)$. The simulation results confirm our theoretical analysis, allowing us to determine c_{eff} and γ for a given noise level. For example, when $\overline{\eta} = 10^{-3}$ with $\Delta_\eta = 5\%$ of $\overline{\eta}$, our hybrid oracle would be applicable up to $n \simeq 27.23$ even in the presence of a phase-flip of $\overline{\chi} = 10^{-2}$, whereas $n \simeq 17.93$ would be the limit in the purely classical case. Equivalently, the hybrid oracle can cover a space up to about $O(e^{1.09\times10^8})$ scale, which is considerably larger than the space, about $O(e^{1.73\times10^5})$ scale, allowed in the classical case. The determined c_{eff} and γ values are listed in Table 1. For details on the methods and results of the numerical analysis, see "Appendix B".

5 Reduction in sample complexity

The quantum advantage described above can be applied to a computational model of learning, the so-called PAC learning model. In this model, a finite set of $h^\star$, say

Table 1 Values of c_{eff} and γ are listed for several cases: $\overline{\chi} = 0$, 10^{-4}, 10^{-3}, and 10^{-2}

$\overline{\chi}$	c_{eff} (cf., $c = 0.5 \times 10^3$)	$\gamma = \frac{c_{\text{eff}}}{c}$
No phase-flip	$0.5 \times 10^{6.23}$	$10^{3.23}$
10^{-4}	$0.5 \times 10^{6.01}$	$10^{3.01}$
10^{-3}	$0.5 \times 10^{5.34}$	$10^{2.34}$
10^{-2}	$0.5 \times 10^{4.40}$	$10^{1.40}$

Here, $\overline{\eta}$ is assumed to be 10^{-3} and $\Delta_\eta = 0.05\overline{\eta}$

$H^\star$, is said to be (ϵ, δ)-PAC learnable and we call the learner (or learning algorithm employed) (ϵ, δ)-PAC learner, if $P[E(h, h^\star) \leq \epsilon] \geq 1 - \delta$ is satisfied for $h^\star \in H^\star$. Here, $E(h, h^\star)$ denotes an arbitrary error function that indicates how h and $h^\star$ are different. Borrowing terms from sampling theory in statistics, ϵ and $1 - \delta$ are defined in terms of the accuracy and confidence, respectively. One beauty of PAC learning is that for any $h^\star \in H^\star$, if a learner is allowed to use a finite size, say $M_b(\epsilon, \delta)$, of samples, then the learner can be a (ϵ, δ)-PAC learner; in other words, we can know whether $H^\star$ is learnable in terms of the required samples, independent of the learning algorithm used. Usually, $M_b(\epsilon, \delta)$ is referred to as sample complexity. For a given set of *noisy* samples $\tilde{T} = \{(\mathbf{x}, m)\}$ where m is either $h^\star(\mathbf{x})$ or $h^\star(\mathbf{x}) \oplus 1$, it is known that $M_b(\epsilon, \delta)$ is at most[3]

$$\frac{2}{\epsilon^2(1 - 2\xi)^2} \ln\left(\frac{2\,|H|}{\delta}\right), \tag{11}$$

where ξ is the imperfection of the oracle, that is, the probability ξ that the oracle provides an incorrect output. For more details about PAC learning, see Refs. [23,30].

Our previous analysis on the improvement of the query success rate can be extended to the reduction in sample complexity by letting $\xi = 1 - \overline{P}_Q(n)$, where $\overline{P}_Q(n)$ is the average query success probability given by $\overline{P}_Q(n) = \frac{1}{2^n} \sum_{\omega=0}^{n} \binom{n}{\omega} P_Q(\omega)$. Note that $\overline{P}_Q(n)$ would decrease as the number of gates (i.e., 2^n) increases, indicating that the oracle's reliability rapidly worsens as the problem size n increases. This is reasonable and shows that the error model is well suited for realistic situations. Here, we introduce the quantity

$$A_Q = \frac{1}{\left(2\overline{P}_Q(n) - 1\right)^2} = \left[\sum_{j=0}^{\infty} \frac{(-1)^j}{j!}\left(\frac{1}{\gamma c}\right)^j \left(\frac{2^j + 1}{2}\right)^n\right]^{-2}, \tag{12}$$

which is consistently proportional to the sample complexity consistently but does not depend on ϵ, δ, and $|H|$. Subsequently, for the sake of comparison, we also consider the case where a purely classical oracle is employed; in this case, the classical counterpart to A_Q is A_C, which is defined using $\overline{P}_C(n) = \frac{1}{2^n} \sum_{\omega=0}^{n} \binom{n}{\omega} P_C(\omega)$ instead of $\overline{P}_Q(n)$.

---

[3] The determination of its optimal (i.e., necessary and sufficient) condition is central and long-standing interest in computational learning theory, but this aspect is outside the scope of our study [37].
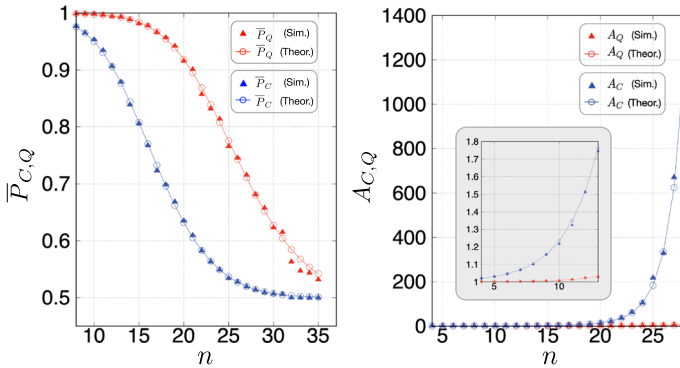
Fig. 2 Numerical plots of $\bar{P}_{C,Q}$ (left) and $A_{C,Q}$ (right) with respect to n. Here, we have used $\bar{\eta} = 10^{-3}$ with $\Delta_\eta = 0.05\bar{\eta}$ and $\bar{\chi} = 10^{-2}$. Theoretical values are also presented for comparison. Refer to the main text for a detailed description

Then, we can show the reduction in the sample complexity from Eq. (10) and Eq. (26) for $A_Q \leq A_C$.

This result tells us that for a large n, our hybrid oracle allows us to define a (ϵ, δ)-PAC learner, whereas a fully-classical oracle cannot. What is remarkable is that the sample complexity reduction is achieved with classical samples without the need to embed them into quantum states, unlike other classical-quantum hybrid schemes. Nevertheless, we note that if n is too large (roughly, when $n \gg 2\log_2 \gamma c$), it is also impractical to define a legitimate PAC learner, even with the quantum learning advantage gained.

To confirm this feature, we perform numerical simulations. $\overline{P}_{C,Q}(n)$ are evaluated by repeating trials for randomly sampled inputs, with each bit x_j ($j = 1, 2, \ldots, n$) in **x** being either 0 or 1 with probability $\frac{1}{2}$, and we examine the range $n = 8$ to $n = 35$. In these simulations, we assume $\bar{\eta} = 10^{-3}$ with $\Delta_\eta = 0.05\bar{\eta}$ and $\bar{\chi} = 10^{-2}$ with $\Delta_\chi = 0.1\bar{\chi}$. In Fig. 2, we plot the dependence of $\overline{P}_{C,Q}$ and $A_{C,Q}$ on n. These results agree well with our theoretical predictions. Detailed descriptions of the method and analysis are provided in "Appendix C".

6 Summary and remarks

We have studied how a tangible quantum advantage can be achieved on the basis of a classical-quantum hybrid architecture distinct from other existing ones. Our approach is appealing, since the proposed hybridization architecture allows classical input data, without requiring (big) classical data to be embedded into a largely superposed quantum state, for example, by implementing QRAM. Instead, the quantum advantage is achieved with a small-sized quantum system (only a single-qubit is required in our case). We applied this approach to a binary classification problem and found that the oracle designed based on our main idea of classical-quantum hybridization, can improve the query success rate. This advantageous feature was attributed to the cance-

lation of the quantum amplitudes associated with the oracle's internal gate errors. Such (say) immunity could lead to a reduction in sample complexity in the PAC learning, facilitating the exploration of a large candidate space with noisy classical samples.

Of the two recent research directions in quantum machine learning—those seeking very strong complexity-theoretic evidence of quantum superiority to be realized in a full-scale quantum computer versus providing prospects for tangible quantum advantages which can be realized with NISQ technologies—our work pertains to the latter. We believe that the presented results are realizable and have potential to facilitate the development of an innovative classical-quantum hybrid technologies.

Appendix A: Detailed calculations of $P_Q(\omega)$

Here, we present the procedure to calculate $P_Q(\omega)$ in Eq. (6) of the main manuscript. We start by analyzing the simple case, i.e., of $\omega = 1$. In particular, we consider an input $\mathbf{x} = x_1 x_2 \cdots x_n$ satisfying $x_{l_1} = 1$ for arbitrary $l_1 \in [1, n]$ and $x_j = 0$ for all $j \neq l_1$. Subsequently, only two gates $\hat{a}_0$ and $\hat{a}_{l_1}$ are activated with $\Omega_{\mathbf{x}} = \{0, l_1\}$. In a purely classical query, $P_C(\omega = 1)$ is given as

$$P_C(\omega = 1) = \left(1 - \eta_{l_1}\right)\left(1 - \eta_0\right) + \eta_{l_1}\eta_0, \tag{13}$$

where η_k is the probability that a bit-flip error will occur at $\hat{a}_k$ ($k \in \{0, l_1\}$). Meanwhile, $P_Q(\omega = 1)$ is calculated as below:
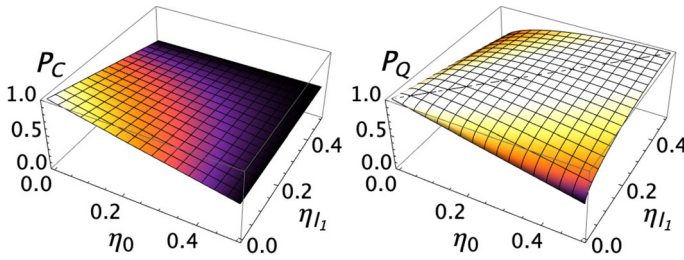
Fig. 3 3D graphs of P_C (left) and P_Q (right) with respect to η_0 and η_{l_1} for $\omega = 1$. The advantage defined in Eq. (16) is observed; P_Q is always larger than P_C. Here, the most remarkable feature is that our hybrid oracle always yields correct results when $P_Q = 1$ provided that $\eta_0 = \eta_{l_1}$

$$
\begin{aligned}
P_Q(\omega = 1) &= \left| \langle h^\star(\mathbf{x}) | \, \hat{\epsilon}_{l_1} \hat{a}_{l_1} \hat{\epsilon}_0 \hat{a}_0 \, | \alpha \rangle \right|^2, \\
&= \left| \langle h^\star(\mathbf{x}) | \left(\sqrt{1 - \eta_{l_1}} \, \hat{\mathbb{1}} \pm i \sqrt{\eta_{l_1}} \hat{\sigma}_x \right) \hat{a}_{l_1} \left(\sqrt{1 - \eta_0} \, \hat{\mathbb{1}} \pm i \sqrt{\eta_0} \hat{\sigma}_x \right) \hat{a}_0 \, | \alpha \rangle \right|^2, \\
&= \Big| \sqrt{1 - \eta_{l_1}} \sqrt{1 - \eta_0} \, \langle h^\star(\mathbf{x}) | \, \hat{a}_{l_1} \hat{a}_0 \, | \alpha \rangle \pm i \sqrt{1 - \eta_{l_1}} \sqrt{\eta_0} \, \langle h^\star(\mathbf{x}) | \, \hat{a}_{l_1} \hat{\sigma}_x \hat{a}_0 \, | \alpha \rangle \\
&\quad \pm i \sqrt{\eta_{l_1}} \sqrt{1 - \eta_0} \, \langle h^\star(\mathbf{x}) | \, \hat{\sigma}_x \hat{a}_{l_1} \hat{a}_0 \, | \alpha \rangle - \sqrt{\eta_{l_1}} \sqrt{\eta_0} \, \langle h^\star(\mathbf{x}) | \, \hat{\sigma}_x \hat{a}_{l_1} \hat{\sigma}_x \hat{a}_0 \, | \alpha \rangle \Big|^2,
\end{aligned}
\tag{14}
$$

where $\hat{\epsilon}_k = \sqrt{1 - \eta_k} \, \hat{\mathbb{1}} \pm i \sqrt{\eta_k} \hat{\sigma}_x$ is the error operation, defined in the main manuscript. Using the properties in Eq. (7) of the main manuscript, i.e., $\hat{\sigma}_x \hat{a}_k = -\hat{a}_k \hat{\sigma}_x$ and $\langle h^\star(\mathbf{x}) | \, \hat{a}_{l_1} \hat{a}_0 \, | \alpha \rangle = 1$, we can evaluate the following:

$$
\begin{cases}
\langle h^\star(\mathbf{x}) | \, \hat{a}_{l_1} \hat{\sigma}_x \hat{a}_0 \, | \alpha \rangle = - \langle h^\star(\mathbf{x}) | \, \hat{a}_{l_1} \hat{a}_0 \hat{\sigma}_x \, | \alpha \rangle = 0, \\
\langle h^\star(\mathbf{x}) | \, \hat{\sigma}_x \hat{a}_{l_1} \hat{a}_0 \, | \alpha \rangle = \langle h^\star(\mathbf{x}) | \, \hat{a}_{l_1} \hat{a}_0 \hat{\sigma}_x \, | \alpha \rangle = 0, \\
\langle h^\star(\mathbf{x}) | \, \hat{\sigma}_x \hat{a}_{l_1} \hat{\sigma}_x \hat{a}_0 \, | \alpha \rangle = - \langle h^\star(\mathbf{x}) | \, \hat{a}_{l_1} \hat{a}_0 \, | \alpha \rangle = -1.
\end{cases}
\tag{15}
$$

Subsequently, using Eq. (15), we can obtain

$$
P_Q(\omega = 1) - P_C(\omega = 1) = \Gamma_{0,l_1},
\tag{16}
$$

where

$$
\Gamma_{0,l_1} = 2\sqrt{1 - \eta_{l_1}} \sqrt{1 - \eta_0} \sqrt{\eta_{l_1}} \sqrt{\eta_0} \geq 0.
\tag{17}
$$

This factor Γ_{0,l_1} is from quantum superposition and clearly indicates the enhancement of the success probability with the condition $\Gamma_{0,l_1} \geq 0$. In Fig. 3, we depict the graphs of $P_{C,Q}$ with respect to η_0 and η_{l_1}. It is noteworthy that our hybrid oracle always yields correct results, i.e., $P_Q = 1$, provided that $\eta_{l_1} = \eta_0$, even though η_{l_1} and η_0 are large. This is the most remarkable feature in our classical–quantum hybrid query.

Subsequently, we consider the case of $\omega = 2$, where a set of four gates, $\hat{a}_0$, $\hat{a}_{l_1}$, $\hat{a}_{l_2}$, and $\hat{a}_{l_3}$, are to be activated with $\Omega_{\mathbf{x}} = \{0, l_1, l_2, l_3\}$. We subsequently calculate

$P_Q(\omega = 2)$ as follows:

$$P_Q(\omega = 2) = \left| \langle h^\star(\mathbf{x}) | \hat{\epsilon}_{l_3} \hat{a}_{l_3} \hat{\epsilon}_{l_2} \hat{a}_{l_2} \hat{\epsilon}_{l_1} \hat{a}_{l_1} \hat{\epsilon}_0 \hat{a}_0 | \alpha \rangle \right|^2. \tag{18}$$

To proceed with the calculation, we introduce an identity $\hat{\mathbb{1}}_{\beta, \beta^\perp} = |\beta\rangle \langle\beta| + |\beta^\perp\rangle \langle\beta^\perp|$, where the state $|\beta\rangle$ ($\beta \in \{0, 1\}$) is defined with the following properties:

$$\left| \langle h^\star(\mathbf{x}) | \hat{a}_{l_3} \hat{a}_{l_2} | \beta \rangle \right|^2 = 1 \text{ and } \left| \langle \beta | \hat{a}_{l_1} \hat{a}_0 | \alpha \rangle \right|^2 = 1. \tag{19}$$

Using a mathematical method of substituting the identity $\hat{\mathbb{1}}_{\beta, \beta^\perp}$ between $\hat{\epsilon}_{l_3} \hat{a}_{l_3} \hat{\epsilon}_{l_2} \hat{a}_{l_2}$ and $\hat{\epsilon}_{l_1} \hat{a}_{l_1} \hat{\epsilon}_0 \hat{a}_0$ in Eq. (18), we can obtain

$$\begin{aligned}
P_Q(\omega = 2) = \Bigg| &\left(\sqrt{1 - \eta_{l_3}} \sqrt{1 - \eta_{l_2}} + \sqrt{\eta_{l_3}} \sqrt{\eta_{l_2}} \right) \left(\sqrt{1 - \eta_{l_1}} \sqrt{1 - \eta_0} + \sqrt{\eta_{l_1}} \sqrt{\eta_0} \right) \\
&- \left(\sqrt{1 - \eta_{l_3}} \sqrt{\eta_{l_2}} - \sqrt{\eta_{l_3}} \sqrt{1 - \eta_{l_2}} \right) \left(\sqrt{1 - \eta_{l_1}} \sqrt{\eta_0} - \sqrt{\eta_{l_1}} \sqrt{1 - \eta_0} \right) \Bigg|^2
\end{aligned} \tag{20}$$

Furthermore, after some algebraic simplifications, we can arrive at

$$\begin{aligned}
P_Q(\omega = 2) - P_C(\omega = 2) = &\, \Gamma_{l_1,0} \left(1 - 2\eta_{l_2} \right) \left(1 - 2\eta_{l_3} \right) + \Gamma_{l_2,l_1} \left(1 - 2\eta_{l_3} \right) \left(1 - 2\eta_0 \right) \\
&+ \Gamma_{l_3,l_2} \left(1 - 2\eta_{l_1} \right) \left(1 - 2\eta_0 \right) + \Gamma_{0,l_3} \left(1 - 2\eta_{l_2} \right) \left(1 - 2\eta_{l_1} \right) \\
&- \Gamma_{l_2,0} \left(1 - 2\eta_{l_3} \right) \left(1 - 2\eta_{l_1} \right) - \Gamma_{l_3,l_1} \left(1 - 2\eta_{l_2} \right) \left(1 - 2\eta_0 \right) \\
&+ 2\Gamma_{l_3,l_2} \Gamma_{l_1,0},
\end{aligned} \tag{21}$$

where

$$\begin{aligned}
P_C(\omega = 2) = &\left(1 - \eta_{l_3} \right) \left(1 - \eta_{l_2} \right) \left(1 - \eta_{l_1} \right) \left(1 - \eta_0 \right) \\
&+ \left(1 - \eta_{l_3} \right) \left(1 - \eta_{l_2} \right) \eta_{l_1} \eta_0 + \left(1 - \eta_{l_3} \right) \eta_{l_2} \left(1 - \eta_{l_1} \right) \eta_0 \\
&+ \eta_{l_3} \left(1 - \eta_{l_2} \right) \left(1 - \eta_{l_1} \right) \eta_0 + \left(1 - \eta_{l_3} \right) \eta_{l_2} \eta_{l_1} \left(1 - \eta_0 \right) \\
&+ \eta_{l_3} \left(1 - \eta_{l_2} \right) \eta_{l_1} \left(1 - \eta_0 \right) + \eta_{l_3} \eta_{l_2} \left(1 - \eta_{l_1} \right) \left(1 - \eta_0 \right) \\
&+ \eta_{l_3} \eta_{l_2} \eta_{l_1} \eta_0.
\end{aligned} \tag{22}$$

Here, $\Gamma_{a,b}$ is defined as $\Gamma_{a,b} = 2\sqrt{1 - \eta_a} \sqrt{1 - \eta_b} \sqrt{\eta_a} \sqrt{\eta_b}$ for $a \neq b \in \Omega_\mathbf{x} = \{0, l_1, l_2, l_3\}$, similarly to Eq. (17). Subsequently, using Eq. (20) and Eq. (21), we demonstrate that the quantum advantage can be achieved with the positive factors $\Gamma_{a,b}$. Note that Eq. (21) could be negative thus exhibiting the disadvantage, e.g., when $\eta_{l_3} = \eta_{l_1} = 0$ or $\eta_{l_2} = \eta_{l_1} = 0$ for all input $\mathbf{x}$. However, the aforementioned situation is not likely to occur in real physical systems. Consistent with the case of $\omega = 1$, we observed that $P_Q(\omega = 2)$ becomes unity when $\eta_{l_3} = \eta_{l_2} = \eta_{l_1} = \eta_0$.

By observing the two cases above, we can infer that the same method, i.e., of introducing the identities, can be used to calculate $P_Q(\omega)$ for arbitrary higher Hamming-weight inputs. The most remarkable construction, i.e., having unity query-success probability with equal error probabilities, can be generalized as well.

Table 2 Detailed values of $\overline{\eta}_{\text{eff}}$ and c_{eff} for each Δ_η

Δ_η	$\overline{\eta}_{\text{eff}}$ (c.f., $\overline{\eta} = 10^{-3}$)	c_{eff} (c.f., $c = 0.5 \times 10^3$)
1% of $\overline{\eta}$	$\simeq 10^{-7.63}$	$\simeq 0.5 \times 10^{7.63}$
5% of $\overline{\eta}$	$\simeq 10^{-6.23}$	$\simeq 0.5 \times 10^{6.23}$
10% of $\overline{\eta}$	$\simeq 10^{-5.60}$	$\simeq 0.5 \times 10^{5.60}$

Therefore, it can be sufficiently concluded that the enhancement in the query-success probability can be achieved for an arbitrary Hamming-weight in our hybrid query.

Appendix B: Numerical analyses with realistic conditions

As mentioned in the main manuscript, in a more realistic situation, the amplitudes related to the errors are not completely canceled out owing to a nonzero Δ_η, and $P_Q(\omega)$ exhibits an analogous form to $P_C(\omega)$ in Eq. (4) of the main manuscript, with an "effective" characteristic constant $c_{\text{eff}} \simeq (2\overline{\eta}_{\text{eff}})^{-1}$. Here, the effective average error $\overline{\eta}_{\text{eff}}$ is expected to be much smaller than c. This feature results in the quantum advantage that does not depend on the degree of $\overline{\eta}$ but only on Δ_η, i.e., how "varying" they are.

To corroborate and extend our theoretical predictions, we perform a numerical analysis. It starts with an input $\mathbf{x}$ of $\omega(\mathbf{x})$. We subsequently evaluate $P_{C,Q}(\mathbf{x})$ by counting the number of "$h^\star(\mathbf{x})$" (e.g.,, "success") and "$h^\star(\mathbf{x}) \oplus 1$" (e.g.,, "failure"), such that $P_{C,Q}(\mathbf{x}) = N_S/(N_S + N_F)$, where N_S and N_F denote the numbers of success and failure, respectively, and $N_S + N_F = 10^5$. Here, we use the Monte-Carlo approach to mimic quantum measurement statistics. This simulation is repeated for different values of η_k (for $k \in \Omega_{\mathbf{x}}$) satisfying $c = (2\overline{\eta})^{-1}$. This condition enables us to analyze the data statistically (i.e., by averaging over the trials) without losing generality, even though in each simulation η_k is changed with different $h^\star$. First, as an extreme but illustrative example, we consider the case of $\Delta_\eta = 0$, i.e., by assuming $\eta_k = \overline{\eta}$ for all possible $k = 1, 2, \ldots, 2^n$. As results, we present the graphs of $P_{C,Q}$ versus ω as dots in Fig. 4a for $\overline{\eta} = 10^{-4}$, 10^{-3}, and 10^{-2}, where each data point of $P_{C,Q}$ is obtained by averaging over $\simeq 10^3$ trials. Here, it is observed that P_C decays fast to $\frac{1}{2}$, indicating good agreement with Eq. (6) of the main manuscript. The data of P_Q are, meanwhile, shown to be unity without depending on the degree of $\overline{\eta}$, as predicted. Next, we consider a realistic situation, assuming that η_k is drawn from a normal distribution $\mathcal{N}(\overline{\eta}, \Delta_\eta)$ for all $k = 1, 2, \ldots, 2^n$ (and hence for $k \in \Omega_{\mathbf{x}}$). Here, we set $\overline{\eta} = 10^{-3}$ with $\Delta_\eta = 1\%$, 5%, and 10% of $\overline{\eta}$. The simulation results are shown in Fig. 4b. For all cases of Δ_η, both P_C and P_Q decay to $\frac{1}{2}$; however, P_Q is much slower. It is also observed that the data of P_Q matched well with Eq. (6) of the main manuscript, thus allowing us to identify the effective characteristic constant c_{eff}. The identified values of c_{eff} and $\overline{\eta}_{\text{eff}}$ are listed in Table 2; they manifest the predicted condition in Eq. (8) of the main manuscript.
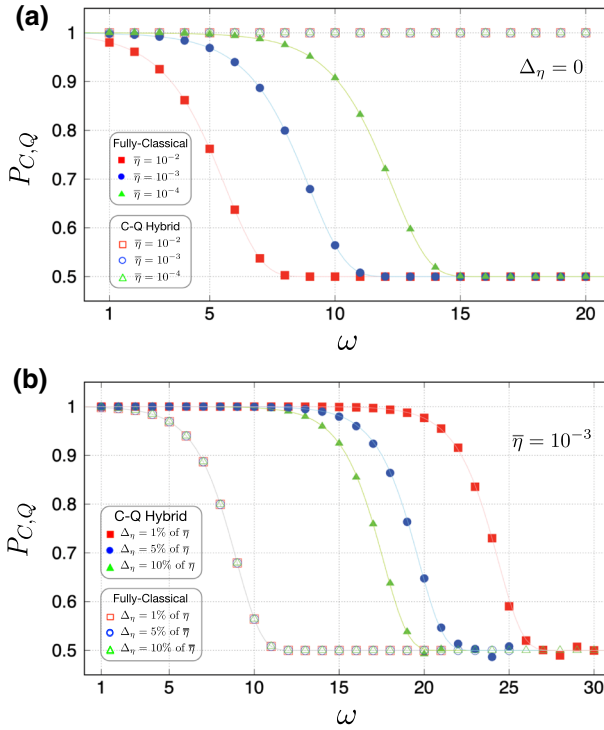
Fig. 4 We plot the graphs of $P_{C,Q}$ versus ω. The simulation is performed for randomly chosen inputs **x** and $h^\star$. In each simulation, $P_{C,Q}$ is evaluated by counting success and failure events over 10^5 queries. One single data point of $P_{C,Q}$ is obtained by averaging $\simeq 10^3$ trials of the simulation. **a** First, we present the simulation data of $P_{C,Q}$ evaluated for $\bar{\eta} = 10^{-4}$, 10^{-3}, and 10^{-2} with $\Delta_\eta = 0$. The results show that P_C rapidly approaches $\frac{1}{2}$ with increasing ω, indicating good agreement with Eq. (6) of our main manuscript (see red, blue, and green solid lines). Meanwhile, P_Q remains unity for all the cases of $\bar{\eta}$, as predicted. **b** Next, we consider the realistic situation, assuming a normal distribution $\mathcal{N}(\bar{\eta}, \Delta_\eta)$. Here, we set $\bar{\eta} = 10^{-3}$ with $\Delta_\eta = 1\%$, 5%, and 10% of $\bar{\eta}$. The data of $P_{C,Q}$ are shown to decay, but P_Q is much slower. In such cases, the data P_Q are well fitted by Eq. (6) of the main manuscript with c_{eff}, indicating that the data agrees well with our theoretical predictions (Color figure online)

Table 3 Detailed values of $\bar{\eta}_{\text{eff}}$ and c_{eff} for each $\overline{\chi}$

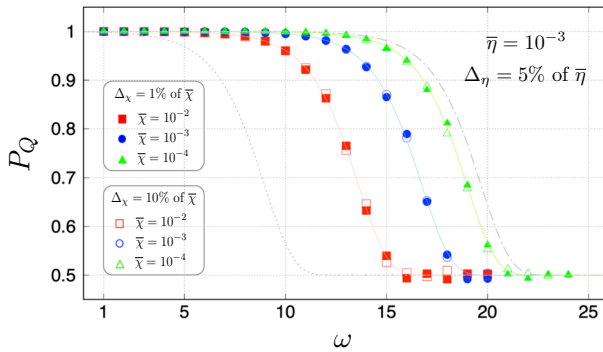| $\overline{\chi}$	$\bar{\eta}_{\text{eff}}$ (c.f., $\bar{\eta} = 10^{-3}$)	c_{eff} (c.f., $c = 0.5 \times 10^3$)
No phase-flip	$\simeq 10^{-6.23}$	$\simeq 0.5 \times 10^{6.23}$
10^{-4}	$\simeq 10^{-6.01}$	$\simeq 0.5 \times 10^{6.01}$
10^{-3}	$\simeq 10^{-5.34}$	$\simeq 0.5 \times 10^{5.34}$
10^{-2}	$\simeq 10^{-4.40}$	$\simeq 0.5 \times 10^{4.40}$

Fig. 5 Graphs of P_Q with respect to ω for $\overline{\chi} = 10^{-4}$, 10^{-3}, and 10^{-2}. Each data point is obtained by averaging over 10^3 simulations. The data fitted well to Eq. (6) in the main manuscript, together with the parameter c_{eff}. The result shows that the quantum advantage becomes less pronounced as $\overline{\chi}$ is increased; however, it is still highly durable. It is noteworthy that the data obtained for $\Delta_\chi = 1\%$ of $\overline{\chi}$ (filled square, circle, and triangle points) and $\Delta_\chi = 10\%$ of $\overline{\chi}$ (empty square, circle, and triangle points) are almost identical (up to the order of 10^{-2}); namely, P_Q is not affected significantly by Δ_χ. The identified $\overline{\eta}_{\text{eff}}$ and c_{eff} are listed in Table 2 (Color figure online)

For a more realistic condition, we consider another type of error, i.e., phase-flip in the assistant qubit that would be crucial for maintaining a higher success rate of the query. In particular, we assume that the phase-flip errors primarily occur when the qubit travels between $\hat{a}_k$ and $\hat{a}_{k+1}$ with a certain probability $\chi_k \leq \frac{1}{2}$. First, when $\Delta_\eta = 0$ (or equivalently, $\eta_k = \overline{\eta}$) for all k, the phase-flip errors do not affect the query process and P_Q becomes unity. In the realistic case, namely of $\Delta_\eta \neq 0$, however, it is predicted that the amplitudes of the bit-flip errors would interfere *disorderly* owing to the phase-flip, and eventually the quantum advantage becomes smaller, as described in our main manuscript. Thus, we perform the simulations and present the data of P_Q in Fig. 5. Here, χ_k is assumed to be drawn from $\mathcal{N}(\overline{\chi}, \Delta_\chi)$ for all $k = 1, 2, \ldots, 2^n$. The simulation data are generated for $\overline{\chi} = 10^{-4}$, 10^{-3} and 10^{-2}. Here, we set $\overline{\eta} = 10^{-3}$ with $\Delta_\eta = 5\%$ of $\overline{\eta}$. The data are well fitted by Eq. (6) of our main manuscript, and c_{eff} are well estimated from the data (see Table 3). As expected, the quantum advantage becomes less pronounced as $\overline{\chi}$ is increased; however, it can still exhibit a higher success rate of the query. It is noteworthy that the data obtained for both $\Delta_\chi = 1\%$ and 10% of $\overline{\chi}$ are almost identical (up to the second digit of a decimal).

Appendix C: Reduction in learning sample complexity in the framework of probably-approximately-correct (PAC) learning

In a probably-approximately-correct (PAC) learning model, a learner (or equivalently, a learning algorithm) samples a finite set of training data $\{(\mathbf{x}_i, h^\star(\mathbf{x}_i))\}$ $(i = 1, 2, \ldots, M)$ by accessing an oracle, aiming at obtaining the best hypothesis h close to $h^\star$ for a given set, e.g., H, of the hypothesis h. Here, $\mathbf{x}_i$ is typically assumed to be drawn uniformly. Subsequently, a learning algorithm is a (ϵ, δ)-PAC learner (under uniform distribution), if the algorithm obtains an ϵ-approximated correct h

with probability $1 - \delta$; more specifically, satisfying

$$P[E(h, h^\star) \leq \epsilon] \geq 1 - \delta, \tag{23}$$

where $E(h, h^\star)$ denotes the error. Here, if h identified by the algorithm agrees with

$$M \geq \frac{1}{\epsilon} \ln \frac{|H|}{\delta} \tag{24}$$

of samples constructed from the oracle, then Eq. (23) holds. Here, $|H| \leq 2^{2^n}$ denotes the cardinality of H. Equation (24) is known as the bound of the sample complexity [23, 24], i.e., it yields the minimum number of training samples to successfully learn $h \in H$ satisfying Eq. (23). Such a sample complexity bound derived from the previous studies can directly be carried over to our scenario; in our classical–quantum hybrid query scheme, the same sample complexity bound exists, because $\mathbf{x}_i$ and $h^\star(\mathbf{x}_i)$ identified by the measurement on $|\psi_{\text{out}}(\mathbf{x}_i)\rangle$ are classical.

However, in the case where the oracle is not perfect, the bound of sample complexity in Eq. (24) is modified as follows: First, we draw a sequence of the training data $\{(\mathbf{x}_1, m_1), (\mathbf{x}_2, m_2), \ldots, (\mathbf{x}_M, m_M)\}$ sampled from our classical–quantum hybrid oracle, where $m_i \in \{h^\star(\mathbf{x}_i), h^\star(\mathbf{x}_i) \oplus 1\}$ denotes the outcome of the measurement performed on $|\psi_{\text{out}}(\mathbf{x}_i)\rangle$. Subsequently, if the sampling is performed with

$$M \geq 2A_Q \ln \left(\frac{2\,|H|}{\delta} \right)^{\frac{1}{\epsilon^2}}, \tag{25}$$

we can verify that Eq. (23) holds for the algorithm that obtains h maximizing $\overline{P}_Q$. In fact, it has been proven that the additional factor A_Q is given as [30]

$$A_Q = \frac{1}{\left(2\overline{P}_Q(n) - 1\right)^2}. \tag{26}$$

It is noteworthy that in the purely classical case, the corresponding factor, e.g., A_C, is given with $\overline{P}_C$ instead of $\overline{P}_Q$. Thus, we can derive the reduction in the sample complexity with the condition $A_Q \leq A_C$ from $\overline{P}_Q \geq \overline{P}_C$. To view this explicitly, we rewrite A_Q in Eq. (26) to a more useful form:

$$A_Q = \left[\frac{1}{2^n} \sum_{\omega=0}^{n} \binom{n}{\omega} e^{-\frac{2\omega}{\gamma c}} \right]^{-2} = \left[\sum_{j=0}^{\infty} \frac{(-1)^j}{j!} \left(\frac{1}{\gamma c} \right)^j \left(\frac{2^j + 1}{2} \right)^n \right]^{-2} \tag{27}$$

This implies that a small increment in the sample complexity bound when n is small increases *abruptly* from near $n \simeq 2 \log_2 \gamma c$. As A_C is characterized by c without γ, we can interpret γ as a quantum learning advantage in the PAC learning framework; i.e., for any large n, we can define a (ϵ, δ)-PAC learner with our hybrid oracle, unlike with a fully classical one. It is noteworthy that if n is excessively large, i.e., when

$n \gg 2 \log_2 \gamma c$, it is impractical to define a legitimate PAC learner even with our hybrid oracle. This result is consistent with the recent theoretical study in Ref. [22]; however, in our case, such a quantum learning advantage is achieved with classical data.

To corroborate and extend our analysis, numerical simulations are performed: For a given n, we prepare a set of inputs $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_M\}$ that is sampled randomly. For the given $\omega(\mathbf{x}_i)$ ($i = 1, 2, \ldots, M$), we evaluate $P_{C,Q}(\omega)$ by counting 10^5 queries and identify their average value, i.e., $\frac{1}{10^5} \sum_{i=1}^{10^5} P_{C,Q}(\omega(\mathbf{x}_i))$. This process is repeated $\simeq 10^3$ times for different input sets to analyze $\overline{P}_{C,Q}(n)$ statistically. The data are generated from $n = 8$ to $n = 35$, assuming that η_k and χ_k ($\forall k$) are drawn from $\mathcal{N}(\overline{\eta}, \Delta_\eta)$ and $\mathcal{N}(\overline{\chi}, \Delta_\chi)$, respectively. Here, we consider $\overline{\eta} = 10^{-3}$ with $\Delta_\eta = 0.05\overline{\eta}$ and $\overline{\chi} = 10^{-2}$ with $\Delta_\chi = 0.1\overline{\chi}$. In each simulation, M is fixed to 100. The obtained data agree well with our theoretical predictions (see Fig. 2a, b in the main manuscript).

References

1. Shor, P.W.: Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer. SIAM Rev. **41**(2), 302–332 (1999)
2. Grover, L.K.: Quantum mechanics helps in searching for a needle in a haystack. Phys. Rev. Lett. **79**, 325 (1997)
3. Harrow, A.W., Hassidim, A., Lloyd, S.: Quantum algorithm for linear systems of equations. Phys. Rev. Lett. **103**, 150502 (2009)
4. Giovannetti, V., Lloyd, S., Maccone, L.: Quantum random access memory. Phys. Rev. Lett. **100**, 160501 (2008)
5. Giovannetti, V., Lloyd, S., Maccone, L.: Architectures for a quantum random access memory. Phys. Rev. A **78**, 052310 (2008)
6. Aaronson, S.: Read the find print. Nat. Phys. **11**, 291 (2015)
7. Arunachalam, S., Gheorghiu, V., Jochym-O'Connor, T., Mosca, M., Srinivasan, P.V.: On the robustness of bucket brigade quantum RAM. New J. Phys. **17**, 123010 (2015)
8. Ciliberto, C., Herbster, M., Ialongo, A.D., Pontil, M., Rocchetto, A., Severini, S., Wossnig, L.: Quantum machine learning: a classical perspective. Proc. R. Soc. A **474**, 20170551 (2018)
9. Preskill, J.: Quantum Computing in the NISQ era and beyond. Quantum **2**, 79 (2018)
10. Arute, F., et al.: Quantum supremacy using a programmable superconducting processor. Nature **574**, 505 (2019)
11. Peruzzo, A., McClean, J.R., Shadbolt, P., Yung, M.H., Zhou, X.Q., Love, P.J., Aspuru-Guzik, A., O'brien, J.L.: A variational eigenvalue solver on a photonic quantum processor. Nat. Commun. **5**, 4213 (2014)
12. McClean, J.R., Romerom, J., Babbush, R., Aspuru-Guzik, A.: The theory of variational hybrid quantum-classical algorithms. New J. Phys. **18**, 023023 (2016)
13. Khoshaman, A., Vinci, W., Denis, B., Andriyash, E., Amin, M.H.: Quantum variational autoencoder. Quant. Sci. Technol. **4**, 014001 (2018)
14. Zhu, D., et al.: Training of quantum circuits on a hybrid quantum computer. Sci. Adv. **5**(10),(2019)
15. Havlíček, V., Córcoles, A.D., Temme, K., Harrow, A.W., Kandala, A., Chow, J.M., Gambetta, J.M.: Supervised learning with quantum-enhanced feature spaces. Nature **567**, 209 (2019)
16. Yoo, S., Bang, J., Lee, C., Lee, J.: A quantum speedup in machine learning: finding a N-bit Boolean function for a classification. New J. Phys. **16**, 103014 (2014)
17. Lee, J.S., Bang, J., Hong, S., Lee, C., Seol, K.H., Lee, J., Lee, K.G.: Experimental demonstration of quantum learning speedup with classical input data. Phys. Rev. A **99**, 012313 (2019)
18. Bang, J., Dutta, A., Lee, S.W., Kim, J.: Optimal usage of quantum random access memory in quantum machine learning. Phys. Rev. A **99**, 012326 (2019)
19. Dunjko, V., Ge, Y., Cirac, J.I.: Computational speedups using small quantum devices. Phys. Rev. Lett. **121**, 250501 (2018)

20. Harrow, A.W.: Small quantum computers and large classical data sets. Preprint arXiv:2004.00026 (2020)
21. Buhrman, H., Newman, I., Rohrig, H., de Wolf, R.: Robust polynomials and quantum algorithms. Theory Comput. Syst. **40**, 379 (2007)
22. Cross, A.W., Smith, G., Smolin, J.A.: Quantum learning robust against noise. Phys. Rev. A **92**, 012327 (2015)
23. Valiant, L.G.: A theory of the learnable. Commun. ACM **27**, 1134 (1984)
24. Langley, P.: Elements of Machine Learning. Morgan Kaufmann (1995)
25. Ambainis, A., Iwama, K., Kawachi, A., Masuda, H., Putra, R.H., Yamashita, S.: Quantum identification of Boolean oracles. In: Annual Symposium on Theoretical Aspects of Computer Science, Springer, pp. 105–116 (2004)
26. Childs, A.M., Kothari, R., Ozols, M., Roetteler, M.: Easy and hard functions for the Boolean hidden shift problem. In: 8th Conference on the Theory of Quantum Computation, Communication and Cryptography (TQC 2013) (Dagstuhl, Germany: Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik) vol. 22 of Leibniz International Proceedings in Informatics (LIPIcs) pp. 50–79 (2013)
27. Arunachalam, S., de Wolf, R.: Guest column: a survey of quantum learning theory. ACM SIGACT News **48**, 41 (2017)
28. Gupta, P., Agrawal, A., Jha, N.K.: An algorithm for synthesis of reversible logic circuits. IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst. **25**, 2317 (2006)
29. Angluin, D., Laird, P.: Queries and concept learning. Mach. Learn. **2**(4), 319 (1988)
30. Angluin, D., Slonim, D.K.: Randomly fallible teachers: learning monotone DNF with an incomplete membership oracle. Mach. Learn. **14**(1), 7 (1994)
31. Bshouty, N.H., Jackson, J.C.: Learning DNF over the uniform distribution using a quantum example oracle. SIAM J. Comput. **28**, 1136 (1998)
32. Toffoli, T.: Reversible computing. International Colloquium on Automata, Languages, and Programming. Springer, pp. 632–644 (1980)
33. Younes, A., Miller, J.F.: Representation of Boolean quantum circuits as Reed-Muller expansions. Int. J. Electron. **91**, 431 (2004)
34. Van Dam, W.: Quantum oracle interrogation: Getting all information for almost half the price. In: Proceedings 39th Annual Symposium on Foundations of Computer Science (Cat. No. 98CB36280) IEEE pp. 362–367 (1998)
35. Debnath, S., Linke, N.M., Figgatt, C., Landsman, K.A., Wright, K., Monroe, C.: Demonstration of a small programmable quantum computer with atomic qubits. Nature **536**, 63 (2016)
36. Eleuch, H., Hilke, M., MacKenzie, R.: Probing Anderson localization using the dynamics of a qubit. Phys. Rev. A **95**, 062114 (2017)
37. Hanneke, S.: The optimal sample complexity of PAC learning. J. Mach. Learn. Res. **17**, 1319 (2016)

Authors and Affiliations

Wooyeong Song[1,7] · Marcin Wieśniak[2,3] · Nana Liu[4,8,9] · Marcin Pawłowski[3] · Jinhyoung Lee[1] · Jaewan Kim[5] · Jeongho Bang[6]

✉ Jinhyoung Lee
   hyoung@hanyang.ac.kr

✉ Jaewan Kim
   jaewan@kias.re.kr

✉ Jeongho Bang
   jbang@etri.re.kr

[1] Department of Physics, Hanyang University, Seoul 04763, Korea

2   Faculty of Mathematics, Physics and Informatics, Institute of Theoretical Physics and Astrophysics, University of Gdańsk, ul. Wita Stwosza 57, 80-306 Gdańsk, Poland

3   International Centre for Theory of Quantum Technologies, University of Gdańsk, ul. Wita Stwosza 63, 80-308 Gdańsk, Poland

4   Institute of Natural Science, Shanghai Jiao Tong University, Shanghai 200240, China

5   School of Computational Sciences, Korea Institute for Advanced Study, Seoul 02455, Korea

6   Electronics and Telecommunications Research Institute, Daejeon 34129, Korea

7   Center for Quantum Information, Korea Institute of Science and Technology (KIST), Seoul 02792, Korea

8   Ministry of Education, Key Laboratory in Scientific and Engineering Computing, Shanghai Jiao Tong University, Shanghai 200240, China

9   University of Michigan-Shanghai Jiao Tong University Joint Institute, Shanghai 200240, China